Modelling of Contempt Context for Content-based Image Retrieval

Huang Jiaqi, Krishna Chandramouli

Division of Enterprise and Cloud Computing,
School of Information Technology and Engineering,
Vellore Institute of Technology University,
Vellore, India 632014.
email: nate.jiaqi@gmail.com, krishna.c@vit.ac.in

Abstract

Relevance feedback is regarded as an invaluable tool to improve the performance of Content-Based Image Retrieval systems, as the techniques embrace user's subjectivity for representing semantic concepts. However, in conventional algorithms relevance feedback is often implemented as a binary classifier, between positive and negative inputs. Therefore, the performance of such approaches is quite limited due to the treatment of both positive and negative samples in a uniformed manner while in reality the negative samples are formed of multiple classes. Addressing the problem of modeling human subjectivity, this paper proposes an adaptive approach based on"contempt context". The contempt context model is implemented as a image ranking algorithm on the Particle Swarm Optimization based image retrieval engine. The user evaluations have been performed to test the proposed model and the results have been reported.

Keywords: Image Retrieval, Contempt Context, Relevance Feedback, Bayesian model, Particle Swarm Optimization, image ranking, user behavior model

1 Introduction

In designing a CBIR system, the first and the most important assumption is that discrimination between relevant and non-relevant items is possible with the available features. Without this condition satisfied relevance feedback is futile. In order to achieve this, many techniques for a straightforward transformation between the topology of the feature space and the semantic characteristics of the items the user wants to retrieve has been reported in the literature. Relevance feedback is often regarded as an invaluable tool to improve CBIR systems, for several reasons. Apart from providing a way to embrace the individuality of users, they are indispensable to overcome the semantic gap between low-level visual features and high-level semantic concepts. By prompting the user for relevance feedback, the initial estimation of relevant documents could be improved to steer the results in the direction the user has in mind. Rather than trying to find better techniques and more enhanced image features in order to improve the performance of what has been referred to as "computer-centric" systems [10], it is more satisfactory to the user to exploit human computer interaction to refine high level queries for representations based on low-level features. This way, the subjectivity of human perception and the user's current context are automatically taken into account as well. Consequently it does not come as a surprise that there exist various techniques of how to make use of relevance feedback in CBIR.

Determining semantic concepts by allowing users to iteratively and interactively refine their queries is a key issue in multimedia content-based retrieval [1]. The Relevance Feedback (RF) loop allows the building of complex queries made out of documents marked as positive and negative examples. From this training set a learning process has to create model of the sought concept from a set of data features to finally provide relevant documents according to the user input. The success of this search strategy relies mainly on the representation spaces within which the data is embedded as well as on the learning algorithm operating in those spaces. These two issues are also intrinsically related to the problem of adequately fusing information arising from different sources.

Addressing the problem of modeling human subjectivity, in this paper we propose an adaptive approach based on "contempt context". The contempt context model is implemented as a image ranking algorithm on the Particle Swarm Optimization based image retrieval engine. The contempt context model is implemented as a image ranking algorithm over the particle swarm optimization based image retrieval engine. The ranking algorithm models contempt context by mapping user inputs, on both positive and negative images to neurons in Self Organizing Maps. The remainder of the paper is organized as follows. In Section 2 an overview of literature review is presented on relevant feedback. The particle swarm optimization based image retrieval framework is presented in 3, followed by the modeling of contempt context in Section 4. Section 5, outlines the experimental results obtained and the comparison between different image retrieval algorithms. The conclusions and future work are presented in Section 6.

2 Related Research

In conventional approaches, the performance of relevance feedback algorithms is still limited, mainly because of the uniform treatment of both positive and negative samples. In reality, the negative feedback samples are likely to be from many different classes (comparing with the query image); therefore, the number of negative feedback samples is normally much greater than the number of positive feedback samples [11]. In this case, an important issue should be considered in relevance feedback, i.e., the im-balanaced positive and negative feedback samples problem. For CBIR RF, data sets are imbalanced and costs are unequal and unknown [6]. This imbalanced data set problem may cause the positive feedback samples to be overwhelmed by the negative instances. Moreover, unlike positive samples, the negative ones are usually highly variable and more difficult to study and utilize. Traditionally, during the RF iteration, the negative samples are always regarded as points from single groups [13] and the differences among different negative samples are lost. The negative samples are sometimes even ignored and only the positive samples are used to address a one-class classification problem [3]. A multi-class based RF algorithm has been proposed in [8] to improve the traditional two-class based (positive and negative) approaches. The algorithm aims to provide additional information to improve the retrieval performance.

Addressing the challenge of effective use of negative samples, in [11] authors propose a new scheme based on a belief that all positive samples are included in a set and the negative samples split into a small number of groups, each one of which has a simple distribution. This means the positive samples form a simple distribution and the negative samples form a complex distribution. Therefore, the negative samples are clustered into groups and then develop a series of sub-classifiers between these negative groups and the unique positive group; finally these sub-classifiers are brought together to make a single classifier. The algorithmic steps presented in the paper are outlined as below.

- Negative samples clustering: The user-labelled negative samples are first clustered into some groups by means of k-nearest neighbors (k-NN); just as in the conventional treatment for the positive samples, it is reasonable to assume that these negative samples, which group together in the feature space, share some common properties and thus can form a single negative class.
- Marginal Convex Machine (MCM): For each of these negative groups, a sub-classifier was build between it and the single positive group. By developing a series of sub-classifiers, the imbalanced problem is avoided to some extent.
- Biased marginal convex machine (BMCM): The above sub-classifiers are then incorporated to a biased marginal convex machine for the positive-negative RF classification. The final classifier not only preserves the merits of each sub-classifier but also has strong generalization ability, i.e., good potential to correctly classify future samples.

In [12], the user's feedback is given in region level and then a set of classifiers is independently trained for each query region. The relevancy of each database image is then obtained by passing the whole image into the trained classifiers and summing up the scores from these classifiers. In [5], an image is represented by a spare vector, in which each element indicates the presence of a certain type of region. The sparse vector representation of the query is then updated by averaging the sparse vectors of positive feedback examples.

3 Image Retrieval based on Particle Swarm Optimization

Particle Swarm Optimization (PSO) is a class of evolutionary computation techniques. It was originally inspired by the social behavior of a flock of birds. The initial study on simulating social behavior of bird flocks and fish schooling were conducted by Reynolds [9] and Heppner [4]. Reynolds was intrigued by the aesthetics of bird flocking choreography while Heppner was interested in discovering the underlying rules that enable a large number of birds to flock synchronously, often changing direction. In PSO, the birds in a flock are symbolically represented as particles. These particles are considered to be flying through a problem space searching for the optimal solution. The location of the particles in a multi-dimensional environment represents a solution to the problem [2].

The motion is attributed to the velocity and position of each particle. Acceleration (or velocity) is weighted with individual parameters governing the acceleration being generated for and . The commonly used PSO versions are global version and local version of PSO. The two versions differ in the update of particles neighborhood, which is generally defined as topologically of knowledge sharing between particles in the swarm. In the local version of PSO, each particles neighborhood includes a limited numbers of particles on its sides, while in the global version of PSO; the topology includes all the particles in the population. The global version of PSO has a fast convergence rate, with potential to converge to the local minimum rather than the global minimum, while the convergence rate of the local version of PSO is slow. The equations governing the velocity and position of each particle are presented in equation 1 and 2.

$$v_{id} = v_{id} + c_1(pbest_i - x_{id}) + c_2(gbest_d - s_{id}) \quad (1)$$

$$x_{id} = x_{id} + v_{id} \quad (2)$$

The trajectory of each individual in the search space is adjusted by dynamically altering the velocity of each particle; according to particles own problem solving experience and the problem solving experience of other particles in the search space.

The image retrieval system, consists of two main subsystems as highlighted in Fig. 1. The first subsystem runs offline and embraces two processing steps. The aim of this step is to extract the different low-level features from the image dataset. The extracted features are stored in the metadata repository. The metadata repository is then further indexed based on the unique ID of image. The second subsystem involves online interaction with the user and comprises a number of processing steps. The second subsystem consists of two online search modules namely "visual search" and "RF system" which are discussed in detail in the following subsections. The reminder of this section will discuss the workflow of the framework.

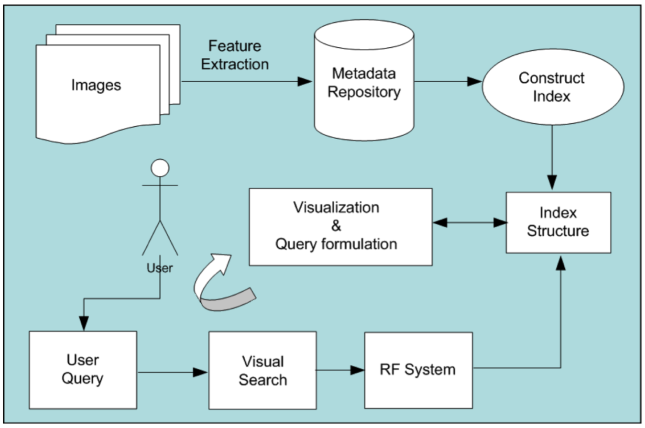

Figure 1. Content-based Image Retrieval framework based on Particle Swarm Optimisation

The interaction is initialized by randomly presenting the user with equal distribution of the database. The user marks only the relevant images from the presented results. The first user interaction inputs are presented to the "visual search module". The visual search module implicitly generates a model for irrelevant model and performs the retrieval. The objective of this step is to infer and predict the user preferences. From the set of results presented from first iteration, the user selects both relevant and irrelevant images and the input is presented to "RF System". The aim of this step is to enhance the inference of the user preferences in order to improve the image retrieval. The user is then iteratively interacts with the system until the user has retrieved all relevant documents or satisfied with the retrieved results.

4 Modelling of Contempt Context

In this paper, the term "Contempt Context" is used to represent the level of dis-satisfaction expressed by the user on the retrieval set of images. Modelling of contempt is achieved in one of two stages. In the first stage, the level of dis-satisfaction over the retrieved image is set is captured, from the number of images the user marks as negative. In the second stage, the concept co-occurence is considered when an image is selected as "relevant" but not positive. The feedback from the user over individual images $x_i$has been modeled with $x_i(label_i, co-occur_i, levelOfContemp_i)$. The $label_i$ represents if the image is considered as positive or negative by the user. The $co-occur_i$ is a boolean value, that is given a high value if the concept is considered to contain semantic concepts that could co-occur with semantic concept of interest. Finally, the $levelOfContempt_i$ is a measure of how much (dis-) interest is shown by the user towards a particular image. The measure $levelOfContempt$ is captured from the amount of user interactions that takes place between the user and the retrieved result.

In Fig. 2, an overview of the contempt context model is pictorially depicted. At a high level, when an user is asked to interact with a CBIR system searching for the concept "boat", different inputs provided by the user are representatively depicted. For example, images with boat concept are marked using blue circles, which correspond to the images satisfying user information need and subsequently, images with water are marked with green. The rational behind such a model, is that boats are likely to appear together with water. Therefore, they share visual characteristics between images depicting boats. Finally the images without any relevance to "boat" concept are marked with "red". For each user feedback, a rank is provided to the media item depending on the interaction. Typically these ranks correspond to how the user ranks between positive, negative and query sets once a retrieval set is presented. In other words, these ranks correspond to the sequence of user-interaction. In modelling contempt context, these additional inputs are also considered in ranking the media items.

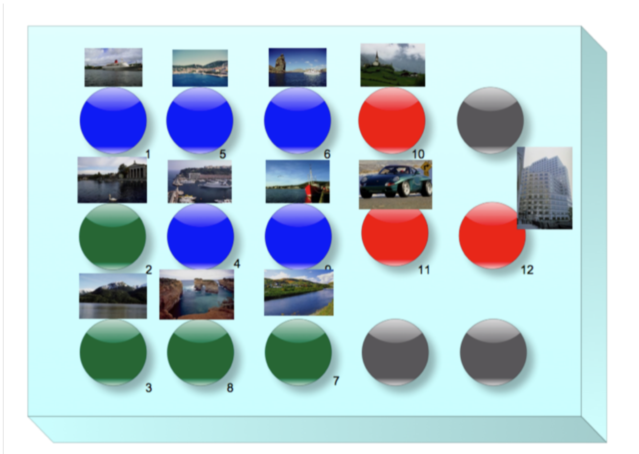

Figure 2. Self Organising Map neurons superimposed by the user feedback

As previously stated, an open challenge of Relevance Feedback techniques can be attributed to how different algorithms process positive and negative samples provided as user feedback. As opposed to previous techniques, where positive set of images are considered to belong to a single category and negative sets are clustered and sub-classifiers are created for each negative cluster with a positive cluster, in this approach each user feedback is considered to be unique. In addition to enabling processing of individual feedback provided by the user, the model also enables ranking of media items with respect to input user sequence. The main advantage of using Self Organizing Maps could be attributed to the fact, that after each neuron is trained online based on the user feedback, the similarity measurement obtained for each media item from the database contains a correspondence to a single winner node. Therefore, depending on the nature of the user feedback used to train the network, its possible to obtain a meaningful ranking of media items from the database. One of the significant advantages of the proposed contempt context includes, the less computationally complexity and no need for a-priori knowledge of the number of user feedback and the nature of user feedback.

5 Experimental Evaluation

The MPEG - 7 visual descriptors namely Color Layout Descriptor (CLD) [7] and Edge Histogram Descriptor (EHD) are extracted for images in the following datasets. The CLD extracts color histograms over image layout. Its similarity measure is a weighted metric with nonlinearly quantized DCT coefficients. The EHD builds on histograms of edges in different directions and scales. Detected edges in a number of directions are used as localized input edge histogram of 80 bins. Its distance is a sum of distances over the original features, as well as global and semi-global histogram values generated by various grouping of local image parts.

The PSO model implemented is a combination of cognitive and social behavior. The structure of the PSO is "fully connected" in which a change in a particle affects the velocity and position of other particles in the group as opposed to partial connectivity, where a change in a particle affects the limited number of neighborhood in the group. Each dimension of the feature set is optimized with 50 particles. The size of the SOM network is pre-fixed with the The database consists of 3500 images annotated with both mid-level features and high-level features. The performance of 4 high-level concepts by 6 different users for both standard-PSO RF and contempt context RF are presented in Figure 3 to Figure 6. On an average the users provided 6 feedback interactions in a specific amount of time. Every user was asked to retrieve as much images as possible within 10 minutes. In average, the contempt based RF model enables users to retrieve more number of images compared to the standard-PSO RF. From the results, it is evident that modeling the user interactions with "contempt context" enhances the performance of the image retrieval system. The increase in performance could be directly attributed to two parameters, namely the concept co-occurence in which semantic concepts that are related to the image query are used as input PSO for optimizing the result set and similarly the use of parameter $levelOfContempt$ to rank the user feedback.

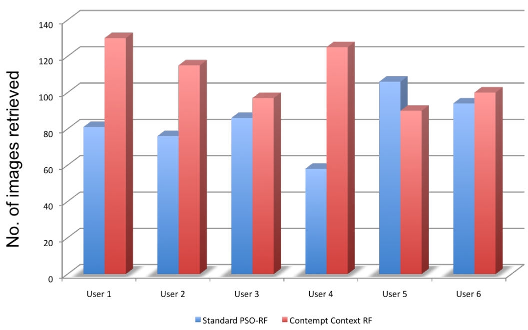

Figure 3. Retrieval Performance for Boat Concept

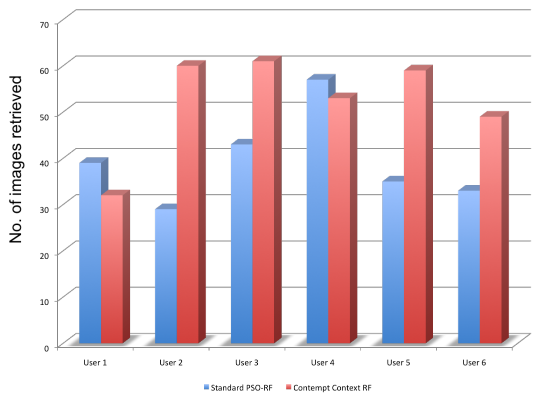

Figure 4. Retrieval Performance for Flower Fields Concept

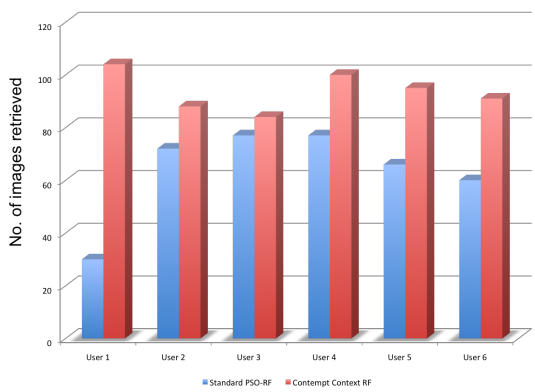

Figure 5. Retrieval Performance for City Street Concept

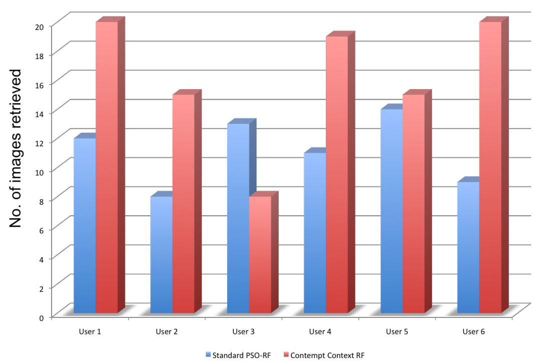

Figure 6. Retrieval Performance for Water Falls Concept

6 Conclusion and Future Work

In this paper, an innovative technique for modeling contempt context based on Self Organizing Maps was proposed along with an extensive evaluation of the technique against similar approaches. The evaluation results from six different users on four semantic concepts highlight the performance of the image retrieval system. The future work will focus on developing Bayesian based models for deriving semantic correlations between multi-layered user interactions. In addition, other research areas to investigate include developing a video based retrieval system as opposed to currently developed keyframe based retrieval system and exploit semantic correlations between concept correlations.

References

[1] E. Bruno, N. Moenne-Loccoz, and S. M. Maillet. Design of multimodal dissimilarity spaces for retrieval of video documents. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 30(5):1520–1533, September 2008.

[2] K. Chandramouli and E. Izquierdo. *Image Retrieval using Particle Swarm Optimization*. CRC Press, 2008.

[3] Y. Chen, X. Zhou, and T. Huang. One-class svm for learning in image retrieval. In *Proc. IEEE International Conference on Image Processing*, pages 815–818, 2001.

[4] U.Grenander F. Heppner. A stochastic non-linear model for coordinated bird flocks. In *The Ubiquity of Chaos. AAAS Publications*, 1990.

[5] F. Jing, M. Li, and H. J. Zhang. An efficient and effective region-based image retrieval framework. *IEEE Trans. on Image Processing*, 13(5):6990709, May 2004.

[6] M. Maloof. Learning when datasets are imbalanced when costs are unequal and unknown. In *Proc. Workshop learning from Imbalanced Data Sets II, held in conjunction with the IEEE Int'l Conf. Machine Learning*, 2003.

[7] B. S. Manjunath, P. Salembier, and T. Sikora. *Introduction to MPEG-7, Multimedia content description interface*. Wiley, New York, 2003.

[8] J. Peng. Multi-class relevance feedback for content based image retrieval. In *Computer Vision and Image Understanding*, pages 42–67, 2003.

[9] C. W. Reynolds. Flocks, herds and schools: A distributed behavior model. In *Computer Graphics*, 1987.

[10] Y. Rui, T. S. Huang, S. Mehrotra, and M. Ortega. Automatic matching tool selection using relevance feedback in mars. In *International Conf. on Visual Information Systems*, 1997.

[11] D. Tao, Z. Li, and S. J. Maybank. Negative samples analysis in relevance feedback. *IEEE Trans. on Knowledge and Data Engineering*, 19(4):568–580, April 2007.

[12] C. Zhang, S. C. Chen, and M. L. Shyu. Multiple objective retrieval for image databases using multiple instance learning and relevance feedback. In *Proc. IEEE Int'l Conf. on Multimedia and Expo*, 2004.

[13] X. Zhou and T. Huang. Relevance feedback fro image retrieval: A comprehensive review. *ACM Multimedia Systems*, 8:536–544, 2003.